

学校编码: 10384

分类号_____密级_____

学 号: 10220120153746

UDC_____

厦 门 大 学

博 士 学 位 论 文

藏 緬 语 演 化 网 络 研 究

Study of Phylogenetic Network in Tibeto-Burman Languages

高 天 俊

指导教师姓名: 邓 晓 华 教 授

专 业 名 称: 语 言 学 及 应 用 语 言 学

论文提交日期: 2 0 1 6 年 0 9 月

论文答辩时间: 2 0 1 6 年 1 1 月

学位授予日期: 2 0 1 6 年 1 2 月

答辩委员会主席: _____

评 阅 人: _____

2016 年 11 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

为检验现有藏缅语研究中所默认采用的谱系树模型的适用性，为更加全面、客观、精确的描述藏缅语的演化模式、演化历史及其谱系关系，解决藏缅语谱系分类研究中的分歧和白语等语言的系属地位争议，本研究将演化生物学中最新发展出来的演化网络理论和演化网生成方法引入藏缅语演化历史研究中，对藏缅语演化网进行了重建和分析。

具体来说，本研究以《藏缅语语音与词汇》中收录的 51 种藏缅语各 1004 条语音词汇材料为基础，开发了藏缅语词汇语音语料库；并在语料库的基础上，开发语音对应规律识别程序，通过概率算法完成了 51 种语言两两之间共 1275 对语言的语音对应规律识别；继而以语音对应规律为依据，开发计算机辅助关系词识别程序并辅以人工检验和判断，识别出 51 种藏缅语言间的 100 核心关系词。最后，本研究以由核心关系词集计算得到的语言距离数据为材料，通过 NeighborNet 演化网生成方法重建了藏缅语演化网，并对可能影响演化网生成结果的几个参数（演化网生成算法、距离算法和词表规模）进行了检验分析。参数检验表明，NeighborNet 是目前最适合于语言演化研究的演化网生成方法；距离算法、词表修正和规模对演化网生成影响不大。

演化网络分析显示：藏缅语的演化历史主要是树状的，分化仍是藏缅语演化历史上占主要地位的演化模式；同时，接触等横向传递模式在藏缅语中也占有不可忽视的地位，在某些语言中，接触可以深入到核心词层面，甚至对其系属地位产生影响。演化网图所反映出的谱系结构与经典分类有较高的契合度，这表明经典的谱系树模型仍适合于描述藏缅语演化历史，经典的分类系统具有较高的准确性。在藏缅语谱系分类的层次划分上，演化网图显示缅彝语群可分为缅、彝、纳西、哈尼四组，这与经典的缅-彝二分法存在差异，提示缅彝语群的层次划分有进一步研究的必要。网图显示白语、土家语在藏缅语谱系系统中的地位独立，提示这两种语言的系属地位或应在汉藏语系层面定义。

本研究首次将演化网络理论和演化网生成方法应用于藏缅语研究中，证明了经典的树状模型在藏缅语研究中的适用性，解决了藏缅语演化历史研究中理论基础不牢的问题。本研究还在全新的网状视角下对经典的谱系分类系统进行了检验

与分析,指示出其有待商榷之处,对白语等系属争议语言的地位给出了新的解读,本研究或可推进这些问题的解决。

关键词: 藏缅语; 演化网; 谱系分类

厦门大学博士论文摘要库

Abstract

In order to test the applicability of the family tree model used in the Tibetan-Burman language study and provide a more comprehensive, objective and accurate description of the evolution model, evolution history and phylogenetic relationship of Tibeto-Burman languages, this study reconstruct and reanalysis the Tibeto-Burman phylogenetic network by the Phylogenetic network reconstruction methods which were newly introduced from Evolutionary Biology.

Specifically, we first develop the Tibetan-Burman phonetic-lexicon corpus based on the phonetic-lexicon materials of 51 Tibetan-Burman languages which were included in *Tibeto-Burman phonology and lexicon* (《藏缅语语音和词汇》). Then we identify the regular sound correspondences in each pair of languages (totally 1275 pairs) with the aid of the Regular Sound-Correspondences Identification Program based on the probabilistic algorithm. After that, the cognates of 100 basic wordlist among 51 languages were identified based on the regular sound-correspondences and some additional rules. Finally, this study reconstructs the Tibetan-Burman evolutionary network by NeighborNet method based on the language distance data calculated from the cognates materials. Additionally, several parameters which may affect the evolution network generation results were tested and analyzed. The parameter test shows that NeighborNet is the most suitable network reconstruct method for language evolution study as well as distance algorithm, wordlist correction and wordlist scale have little effect on evolution network reconstruction.

Phylogenetic network figures show that the history of Tibeto-Burman languages is mainly tree-like, split is still the dominant mechanism in the history. Meanwhile, the horizontal transmission plays a significant role in the history of Tibeto-Burman languages. In some particular languages, the language contact may deep into the core vocabularies and may impact the phylogenetic position of related languages. The phylogenetic structure shows in network figures and classic phylogenetic classification are very similar, this shows that the classical family tree model is still suitable for describing the evolution history of Tibeto-Burman languages. On the

issue of classification hierarchy, network figures show that the Lolo-Burmese group can be divided into 4 subgroups: Burmese, Loloish, Naxi and Hani, which is different from the classic classification. This difference implies that it is necessary to reanalysis the hierarchy of Lolo-Burmese group. Furthermore, the independent status of Bai and Tujia imply that the positions of these two languages should be defined at the level of Sino-Tibetan family.

In this study, the evolution network theory and the evolution network reconstruction methods are applied to the study of Tibeto-Burman language for the first time, which proves the applicability of the classical family tree model in Tibeto-Burman studies and solves the theoretical foundation problem of Tibet-Burman history studies. This study also examines and analyzes the classical classification system under the new network perspective, and gives a new interpretation of the phylogenetic position of some disputed languages such as Bai. This study may promote the resolution of these problems.

Key Words: Tibeto-Burman Languages; Phylogenetic network; Phylogenetic classification

目录

摘要	I
Abstract	III
目录	V
Table of Contents	VII
表格目录	IX
图示目录	X
1 绪论	1
1.1 中国境内藏缅语概况	1
1.2 藏缅语分类研究概况	5
1.3 藏缅语谱系分类研究中的问题	12
1.4 藏缅语谱系分类问题的原因	15
1.5 藏缅语演化分类研究的新探索	17
1.6 本课题研究的内容和目的	19
1.7 本课题研究的材料和方法	20
1.8 本课题的研究思路和本文的结构设计	20
2 语言谱系分类方法与藏缅语定量分类	23
2.1 语言分类的类别	23
2.2 语言谱系分类的标准与方法	24
2.3 定性分类法与定量分类法	29
2.4 定量分类方法的原理及在语言分类上的应用	33
2.5 词源统计法与定量分子种系发生法的结合	42
3 语言演化网络模型与演化网生成方法	45
3.1 树状演化模型	45
3.2 汉藏语系演化历史的复杂性	50
3.3 演化网生成方法	52
4 藏缅语核心词表	61
4.1 核心词表	61
4.2 藏缅语 100 核心词修正方案	67
4.3 藏缅语谱系网络研究百词表	70
5 藏缅语语音对应规律和关系词识别	73
5.1 关系词识别研究的方向	73
5.2 历史语言学中的统计学规律	74
5.3 语音对应规律和关系词自动识别研究概述	75
5.4 计算机辅助藏缅语语音对应规律和关系词识别	83
5.5 关系词识别效果的比较检验	93
6 藏缅语演化网重建与分析	95
6.1 藏缅语演化网络重建的材料、方法与步骤	95
6.2 藏缅语演化网重建的方法与材料的变量分析	98
6.3 藏缅语演化网络分析	106
7 结语	117
7.1 本研究的主要内容、创新点及研究意义	117
7.2 本研究的不足及展望	118

附录一：51 种藏缅语编号.....	121
附录二：藏缅语核心关系词表.....	123
参考文献.....	174
中文文献.....	174
英文文献.....	179
致谢	193

厦门大学博士论文摘要库

Table of Contents

Abstract	III
Table of Contents	VII
List of Tables	IX
List of Figures	X
1 Introduction	1
1.1 Tibeto-Burman Languages in China	1
1.2 Overview of Tibeto-Burman Classification study	5
1.3 Problems of Tibeto-Burman Classification	12
1.4 Causes of Tibeto-Burman Classification problems	15
1.5 Exploring in Tibeto-Burman Classification study	17
1.6 Content and purpose of this study	19
1.7 Materials and methods of this study	20
1.8 The research ideas and structure of this thesis	20
2 Phylogenetic classification methods and quantitative classification of Tibeto-Burman languages	23
2.1 Type of language classification	23
2.2 Criteria and methods of language phylogenetic classification	24
2.3 Qualitative and quantitative methods in language classification	29
2.4 Principle of quantitative classification method and its application in the language classification	33
2.5 The combination of Lexicostatistics and quantitative molecular phylogenetic classification methods	42
3 Phylogenetic Network model and phylogenetic reconstruction methods of language evolution	45
3.1 Phylogenetic tree model	45
3.2 The complexity of the history of Sino-Tibetan languages	50
3.3 Phylogenetic network reconstruction methods	52
4 Basic word lists in Tibeto-Burman languages	61
4.1 Basic word lists	61
4.2 Revision of Tibeto-Burman 100 basic wordlist	67
4.3 100 basic wordlist for the study of phylogenetic network study in Tibeto-Burman languages	70
5 Sound correspondences and cognates Identification in Tibeto-Burman languages	73
5.1 The direction of cognates identification study	73
5.2 Statistical law in historical linguistics	74
5.3 Overview of automatical sound correspondences and cognates Identification	75
5.4 Computer assisted regular sound correspondences and cognates identification on Tibeto-Berman languages	83
5.5 Evaluation of cognates identification	93
6 Reconstruction and analysis of Tibeto-Burman phylogenetic network	95
6.1 Materials, methods and procedures of the phylogenetic network reconstruction in Tibeto-Burman languages	95

6.2	Variable analysis on the methods and materials of phylogenetic network reconstruction.....	98
6.3	Analysis on Tibeto-Burman phylogenetic networks	106
7	Conclusion	117
7.1	The main contents and innovations of this study	117
7.2	Insufficient and prospects of this study.....	118
	Appendix I: Index of 51 Tibeto-Burman Languages	121
	Appendix II: Cognates of Tibeto Burman Languages.....	123
	References.....	174
	Acknowledgments.....	193

表格目录

表 1-1: 藏缅语言的使用人数	2
表 2-1: 词源统计法与语言年代学的基本假设差异.....	28
表 2-2: 词源统计法与语言年代学的操作步骤差异.....	29
表 2-3: 郑锦全(1988)对汉语方言词汇的特征编码.....	38
表 2-4: 多元编码与二元编码的转换.....	39
表 4-1: Swadesh (1955) 200/100 核心词表.....	63
表 4-2: Swadesh (1971) 100 词表.....	64
表 4-3: 藏缅语中“羊”、“山羊”和“绵羊”的关系.....	69
表 4-4: 缺词及重复词项处理方案.....	70
表 4-5: 新增词条及其来源	70
表 4-6: 藏缅语谱系网络研究百词表.....	71
表 5-1: 两种不同的语音对齐	82
表 5-2: 疑似源材料错误表	84
表 5-3: 词表中可能包含有重复语素的词.....	85
表 5-4: 音节对齐表数据结构示例.....	87
表 5-5: 语音对应规律存储数据结构示例.....	89
表 6-1: 51 种藏缅语汉明距离矩阵	96
表 6-2: 图 6-2 所对应的 Delta Score 和 Q-residual Score	107
表 6-3: 藏缅语的分类层次	110

图示目录

图 1-1: 《中国语言地图集》藏缅语族语言地图.....	3
图 1-2: 邓晓华、王士元 (2003a) 藏缅语 100 核心词同源百分比.....	4
图 1-3: 科诺的藏缅语分类	5
图 1-4: 罗常培、傅懋勤的藏缅语分类.....	6
图 1-5: 谢飞的汉藏语分类	7
图 1-6: 白保罗的汉藏语分类系统.....	8
图 1-7: 戴庆厦的藏缅语分类	9
图 1-8: 马提索夫的汉藏语分类	10
图 1-9: 《中国的语言》中的藏缅语分类.....	11
图 1-10: 邓晓华、王士元的藏缅语定量谱系分类.....	12
图 1-11: 49 种藏缅语言分布图	22
图 2-1: 3 个分类对象的不同演化路径.....	34
图 2-2: 语言定量分类法的操作步骤.....	37
图 2-3: 邓晓华、王士元 (2003c) 使用的距离转换函数图.....	41
图 3-1: 分化与杂交的不同演化模式.....	47
图 3-2: 迪克森关于系族分枝和区域影响的假想图.....	50
图 3-3: 由特征数据得到无根树	53
图 3-4: 由特征数据生成演化网	54
图 3-5: Bryant et al. (2005) 84 种印欧语 NeighborNet 网图	56
图 3-6: McMahon 等为 26 种澳大利亚语言生成的网图	57
图 3-7: Harmed and Wang (2006) 为汉语方言生成的网图	58
图 5-1: 两种不同的自动识别程序流程图.....	79
图 5-2: 期望次数为 1.818 时出现指定次数对当的概率.....	89
图 5-3: 藏缅语关系词辅助识别系统功能概览.....	92
图 5-4: 基于 100 核心关系词的藏缅语谱系树图.....	94
图 6-1: 藏缅语 100 核心关系词汉明距离 Split Decomposition 网图	99
图 6-2: 藏缅语 100 核心关系词汉明距离 NeighborNet 演化网图.....	100
图 6-3: 藏缅语 100 关系词负对数距离 NeighborNet 演化网图.....	101
图 6-4: 藏缅语 93 核心关系词 NeighborNet 演化网图.....	103
图 6-5: 藏缅语 35 核心关系词 NeighborNet 演化网图.....	105
图 6-6: 藏缅语 Neighbor-joining 树图	108
图 6-7: 网图所显示出的藏缅语南北分群.....	111
图 6-8: 藏缅语演化网与谱系分类的地理分布.....	116

1 绪论

本课题利用大规模汉藏语系语言基本词汇数据库,运用演化生物学中最新发展出来的种系发生理论和演化网生成方法,研究中国境内藏缅语诸语言的谱系关系和演化网络题。

1.1 中国境内藏缅语概况

语言的谱系分类是研究语言演化历史成果的一种展示方式,也是历史语言学关注的重点和焦点之一。自历史语言学诞生以来,印欧语学者在印欧语的谱系分类研究上取得了辉煌的成果,已经为印欧语建立了完整、可信的谱系分类系统。而在汉藏语研究方面,自李方桂(Li 1937)提出汉藏语的系属意见以来,国内外学界对汉藏语的分类众说纷纭,各种观点交锋激烈。

汉藏语谱系分类的争议是其独特的演化历史造成的:汉藏语系的演化历史与印欧语,特别是南岛语存在着显著的差异。印欧语是在相对独立的环境下从原始印欧语发展而来的,其分化历史上未受其他语系太大影响,语言之间的语音对应规律线索受破坏较少,线索仍比较清晰,因此能较确切的重建其演化的历史,并使用谱系树对其历史分类进行相当好的描述。南岛语的演化历史线索则更为清晰:南岛语族群在太平洋群岛上迁徙、扩散的路径可以用“快车模型(Express Train)”来描述,即操(古)南岛语的人从一个岛扩散到另一个岛,他们到达新的陆地后,很少和故乡联系,从而形成一种链状的迁徙路径。在南岛族群迁徙的过程中,语言也随之扩散、演化,由于各岛屿之间接触很少,横向的语言扩散在南岛语系中影响很小,这使得南岛语系的演化历史体现出较纯粹的以纵向传递为主的特点,也使得其历史非常适合于谱系树模式。

汉藏语系诸语言所处的地理环境以及其演化历程与欧洲、大洋洲有着很大的差异。一直以来,东亚、东南亚地区都有诸多民族杂处混居,语言接触频繁,语言特征的横向扩散与纵向遗传混杂在一起,语言的分化式演化(树状模式)和融合式演化(波浪模式、语言联盟)这两种模式相互纠葛、难以区分。另一方面,本区域内部的各民族语言近两千年来一直受到强大的汉文化的影响,大量的汉语成分与特征渗入相关语言中,语言的接触甚至换用频繁发生,极大的改变了该地区的语言关系面貌。这些独特的演化模式使得经典的历史比较法以及谱系树模式

在进行汉藏语历史研究时难免遇到困难，谱系分类的争议也由此而来。

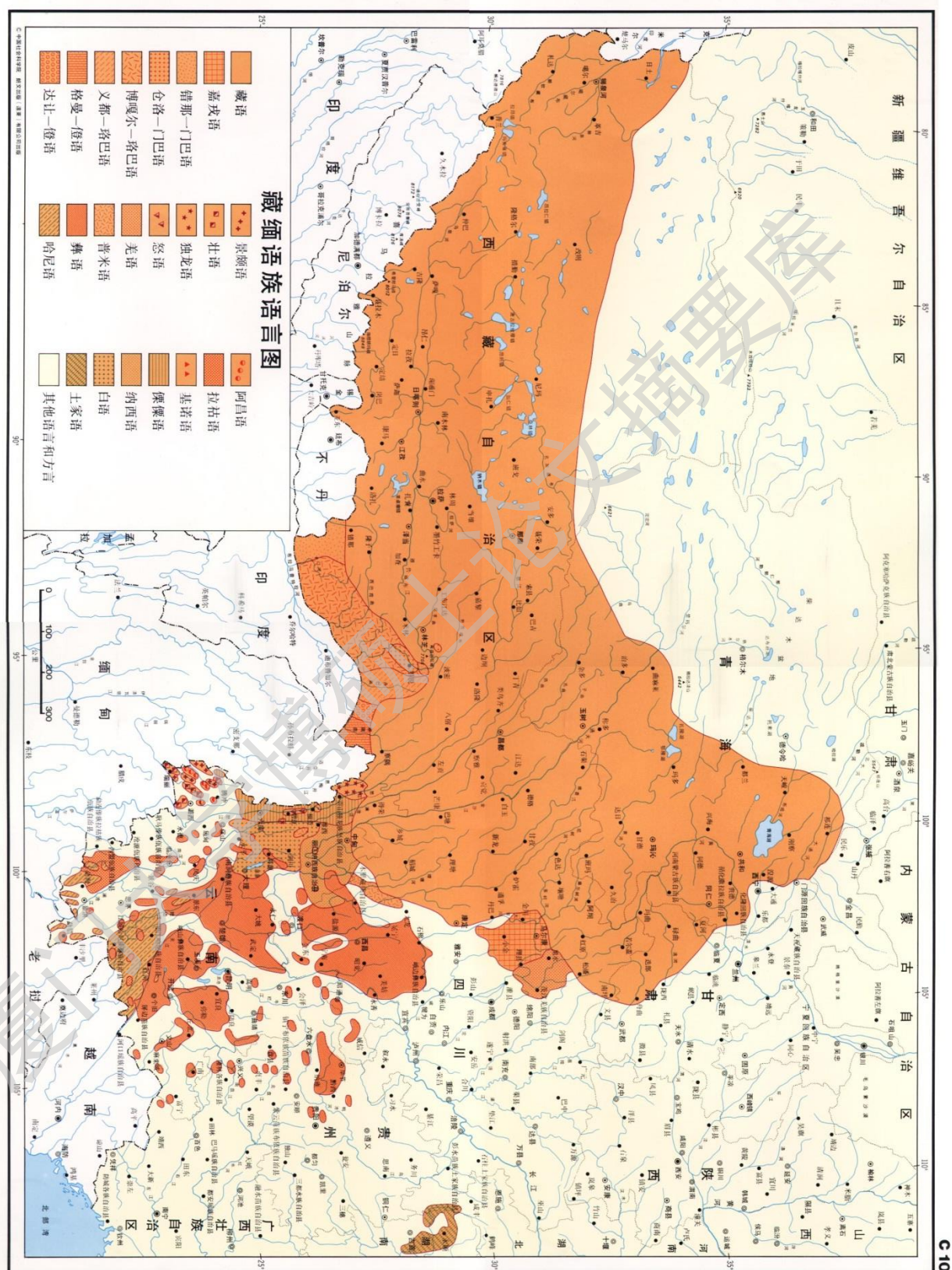
具体来看，在语系层面，苗瑶、壮侗语的归属及汉-南岛语的关系问题至今仍争议难平（Li 1937, Benedict 1942, 罗常培、傅懋绩 1954, Benedict 1972, 邢公畹 1991, 郑张尚芳 1995, 潘悟云 1995）。在语族内部分类问题上，李方桂的汉藏语体系中，在壮侗语（罗常培、傅懋绩 1954, 梁敏、张均如 1996）和苗瑶语（中国科学院少数民族语言研究所编 1959, 王辅世, 毛宗武著 1995）的谱系分类问题上，除壮侗语的临高话和苗瑶语的畲语争议较大外，其整体分类已比较清楚。而相对而言，藏缅语的情况则要复杂得多。

首先，藏缅语分布地域广阔、涉及语言众多。藏缅语族是汉藏语系中分布最广、语言数量最多的一支。其分布区域遍及中国、缅甸、泰国、越南、巴基斯坦、孟加拉、老挝、印度、尼泊尔、不丹等在内的多个国家（戴庆厦等 1989），已知有至少约 250 种语言^①（Matisoff 1991），其中使用人数超过 100 万的语言就有 9 个（表 1-1）。在中国境内共有 46 种藏缅语言（孙宏开等主编 2007），分布在西藏、青海、甘肃、四川、云南、贵州和湖南诸省在内的广大地区（见图 1-1），包括藏、门巴、珞巴、景颇、独龙、怒、羌、普米、彝、哈尼、阿昌、拉祜、基诺、傈僳、纳西、白和土家在内的 17 个民族都使用藏缅语^②。

表 1-1：藏缅语言的使用人数^③

使用人数（单位：人）	语言数
超过 1,000,000	9
500,000 - 999,000	12
250,000 - 499,000	11
100,000 - 249,000	16
50,000 - 99,000	16
25,000-49,000	27
10,000 - 24,000	44
不到 10,000	123

① 根据语言和方言区分标准的宽严不同，不同的统计数字可能有差异。
② 目前白语的系属问题尚有争议，这里暂时将其列入藏缅语中讨论。
③ 材料来源：Matisoff J A 1991. Sino-Tibetan Linguistics: Present State and Future Prospects[J]. Annual Review of Anthropology,20: 469-504.

图 1-1:《中国语言地图集》藏缅语族语言地图^①

^① 材料来源: 中国社会科学院, 澳大利亚人文科学院 1987. 中国语言地图集 [M]. 香港: 朗文出版(远东)有限公司

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.